

FIG. 1

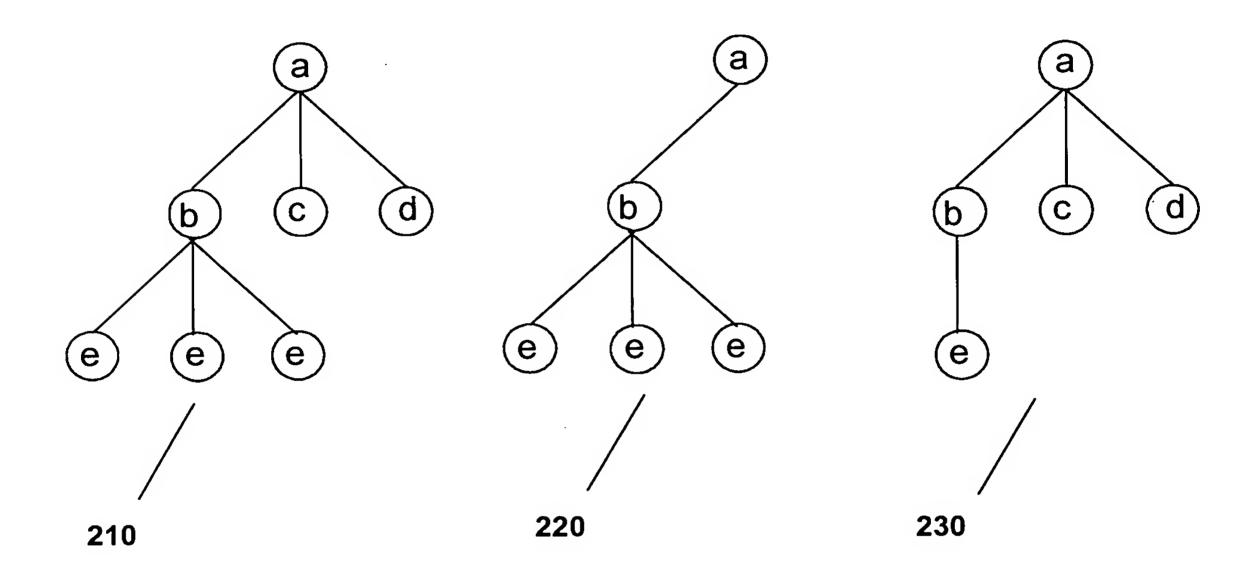


FIG. 2

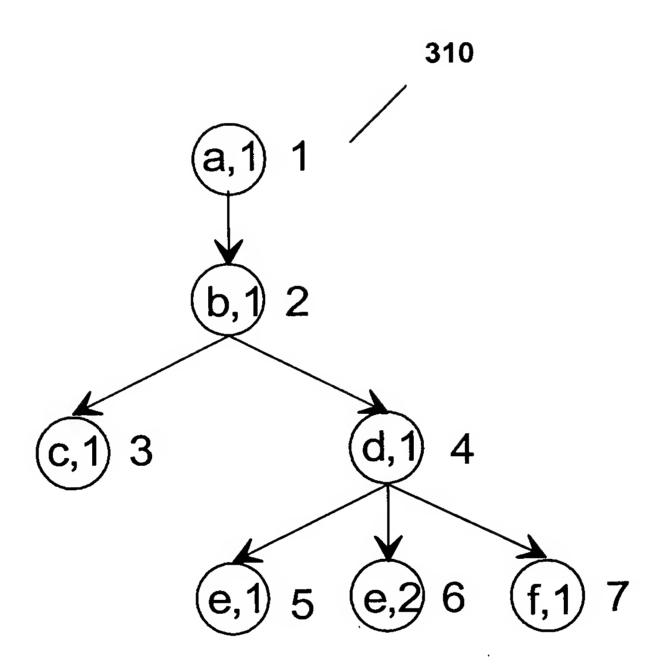


FIG. 3

Build a dictionary Dict based on bag of Paths or bag of XPaths model

Represent d_i and d_l as N dimensional vectors $[d_{il}, d_{i2},.., d_{iN}]$ and $[d_{jl}, d_{j2},.., d_{jN}]$

Compute the similarity measure for documents d_i and d_l

FIG. 4

430

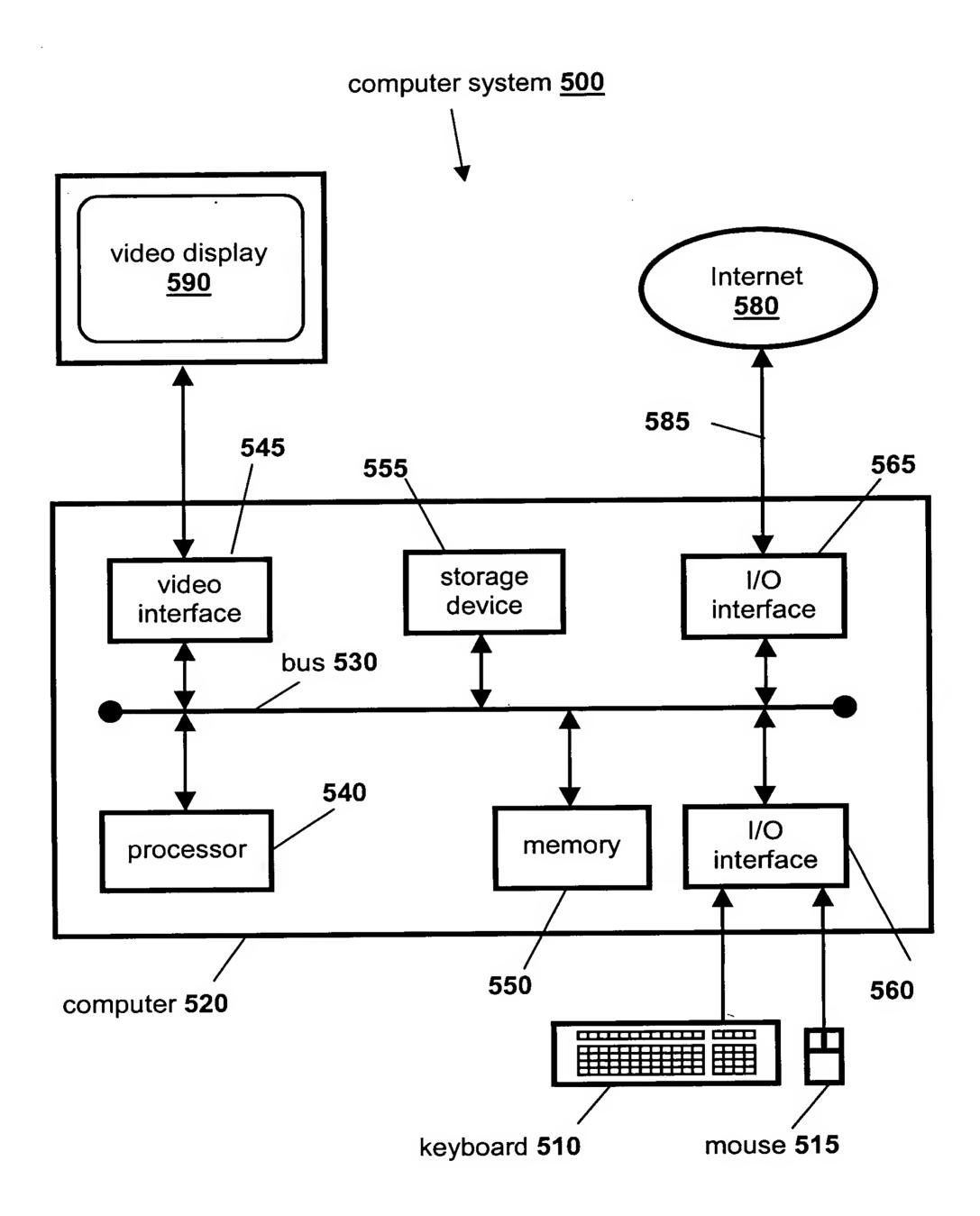


FIG. 5

2003.06.26